Lasso Regression

Thursday, June 10, 2021 6:42 AM

L1 Regularization | overfitting → L2 reg

$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda \|w\|^2$$

↳ $\lambda (w_1^2 + w_2^2 + \dots + w_n^2)$

→ 0 $w_1 \to w_n \to$ coeff

↳ Lasso L1 norm → $[|w_1| + |w_2| + |w_3| + \dots + |w_n|]$

$$L = MSE + \lambda \|w\|$$

→ code implement (sklearn)

→ key point →

→ difference Ridge Vs Lasso

Lasso Regression | Intuition and Code Sample | Regularized Linear Models

Home Insert Draw View Help

Shapes Ink to Shape Ink to Text

100 Days of ML

Day38-Missing Indi...
Day39 - KNN Impu...
Day40 - Iterative I...
Day 41 - Outliers in...
Day 42 - Outlier De...
Day 43 - Outlier de...
Day44 - Outlier De...
Day 45 - Feature C...
Day 46 - Curse of...
Day 47 - PCA
Day 48 - Simple Li...
Day 49 - Regressio...
Day 50 - Multiple L...
Day 51 - Gradient...
Day 52 - Types of...
Day 53 - Polynomi...
Day 54 - Bias Varia...
Day 55 - Ridge Reg...
Day 56 - Lasso Reg...

Lasso Regression

alpha

↳ Lasso

L1 norm → $|w_1| + |w_2| + \dots$

$L = MSE + \lambda \|w\|$

→ code implement (sklearn)

→ key point →

→ Difference Ridge Vs Lasso

underfitting

$\lambda = 0 \rightsquigarrow$ Linear

overfitting

degree

x_1 x_2 (x) o o x_n (y)

0.5 (0.2)

ridge

lasso → $\lambda \rightarrow \uparrow$

dim <

feature selection

0 decrease

lasso

dim < decrease

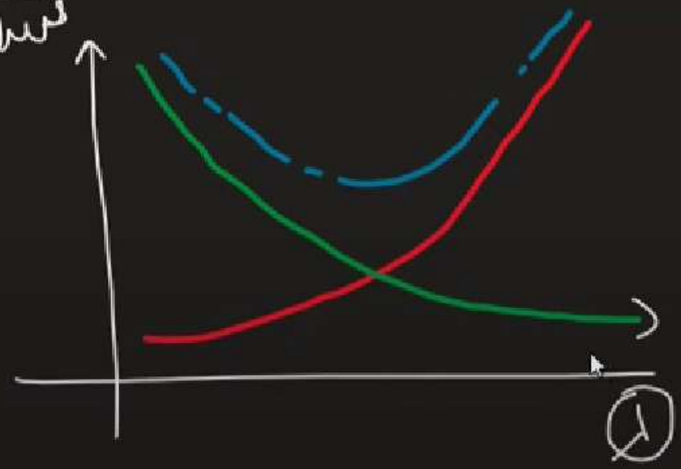
$\lambda \uparrow$

overfitting ↓ Bias ↑

Variance ↓

Variance ↑ Bias ↓

value



Red → bias

green – Variance

blue – total error

0.0001	-9.071288	-205.337332	516.780313	340.539730	-888.652320	555.952271	150.585260	125.453044	858.639860	52.379002
0.0010	-8.264924	-204.213177	517.641106	339.751339	-826.653342	508.609613	120.899583	113.924518	836.314382	52.011583
0.0100	-1.361404	-192.944226	526.348511	332.649058	-430.205495	191.277876	-44.048113	68.990747	688.384976	47.939528
0.1000	0.000000	-113.976046	526.737112	292.635423	-82.691928	-0.000000	-152.691332	0.000000	551.077200	7.169852
1.0000	0.000000	0.000000	363.882636	27.278420	0.000000	0.000000	-0.000000	0.000000	336.135971	0.000000
10.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-0.000000	0.000000	0.000000	0.000000
100.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-0.000000	0.000000	0.000000	0.000000
1000.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-0.000000	0.000000	0.000000	0.000000
10000.0000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	-0.000000	0.000000	0.000000	0.000000

$$m = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2 + \lambda}$$

Simple

x | y → $y = mx + b$

$b = \bar{y} - m\bar{x}$

$\bar{y} \rightarrow \text{mean}(y)$

$\bar{x} \rightarrow \text{mean}(x)$

$$m = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Why Lasso Regression creates sparsity?

$$b = \bar{y} - m\bar{x}$$

$\bar{y} \rightarrow \text{mean}(y)$

$\bar{x} \rightarrow \text{mean}(x)$

$$m = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$b = \bar{y} - m\bar{x}$ $\quad m = ?$

$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda |m|$$

$$\frac{d}{dm} \sum_{i=1}^{n} (y_i - mx_i - \bar{y} + m\bar{x})^2 + 2\lambda m$$

$$\frac{dL}{dm} = \sum_{i=1}^{n} (y_i - mx_i - \bar{y} + m\bar{x})^2 + 2\lambda |m| = 2 \sum (y_i - mx_i - \bar{y} + m\bar{x})(-x_i + \bar{x}) + 2\lambda = 0$$

Why Lasso Regression creates sparsity?

$$L = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 + \lambda |m|$$

$$\frac{d}{dm} \sum_{i=1}^{n} (y_i - mx_i - \bar{y} + m\bar{x})^2 + 2\lambda m$$

$$\frac{dL}{dm} = \sum_{i=1}^{n} (y_i - mx_i - \bar{y} + m\bar{x})^2 + 2\lambda |m| = 2\sum (y_i - mx_i - \bar{y} + m\bar{x})(-x_i + \bar{x}) + 2\lambda = 0$$

$$-2\sum [(y_i - \bar{y}) - m(x_i - \bar{x})](x_i - \bar{x}) + 2\lambda = 0$$

$$-\sum [(y_i - \bar{y})(x_i - \bar{x}) - m(x_i - \bar{x})^2] + \lambda = 0$$

$$-\sum (y_i - \bar{y})(x_i - \bar{x}) + m\sum (x_i - \bar{x})^2 + \lambda = 0$$

$$m\sum (x_i - \bar{x})^2 = \sum (y_i - \bar{y})(x_i - \bar{x}) - \lambda$$

$$m = \frac{\sum (y_i - \bar{y})(x_i - \bar{x}) - \lambda}{\sum (x_i - \bar{x})^2}$$

Lasso coeff sparsity

for $m > 0$

$$m = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})^2 - \lambda}{\sum (x_i - \bar{x})^2}$$

for $m = 0$

$$m = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

for $m < 0$

$$m = \frac{\sum (y_i - \bar{y})(x_i - \bar{x}) + \lambda}{\sum (x_i - \bar{x})^2}$$

Lasso

for $m > 0$

$$m = \frac{\sum (y_i - \bar{y})(x_i - \bar{x}) - \lambda}{\sum (x_i - \bar{x})^2}$$

$\lambda > 0$

for $m = 0$

$$m = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2}$$

for $m < 0$

$$m = \frac{\sum (y_i - \bar{y})(x_i - \bar{x}) + \lambda}{\sum (x_i - \bar{x})^2}$$

$m \rightarrow$ 2, $\frac{9}{5}$, 1, 0, -1, .5, 0

$\lambda \uparrow$

$$m = \frac{YX - \lambda}{X^2}$$

$Y_{X=100}$

$X^2 = 50$

$\lambda = 100$

$m = 0$

$\lambda > 100$

$m = -1$

$$m = \frac{YX + \lambda}{X^2} = \frac{100 + \lambda}{50}$$

$$= \frac{100 + 150}{50} =$$

$m = 5$

$$m = \frac{100 - \lambda}{50}$$

$\lambda = 0 \quad m = 2$

$\lambda = 10 \quad m = \frac{9}{5}$

$\lambda = 50 \quad m = 1$

150

150

$m = \frac{100 - \lambda}{50}$ $\quad m = 2$ $\quad m = \frac{9}{5}$

$\lambda = 50 \quad m = 1$

$m = -1$ $\qquad 150$

$= \frac{100 + 150}{50} =$

$m = 5$

$m < 0$ $\qquad m = \frac{\sum (y_i - \bar{y})(x_i - \bar{x}) + \lambda}{\sum (x_i - \bar{x})^2}$ $\qquad m = \frac{-100 - \lambda}{50}$

$\lambda > 0$

$= \frac{-100 - 150}{50} = -5$

$m = \frac{-100 + \lambda}{50}$ $\qquad \lambda = 0 \quad m = -2$

$\lambda = 50 \quad m = -1$

$\lambda = 100 \quad m = 0 \rightarrow 1$

$\lambda = 150$

$\lambda = 150, \; m = -5$

1) $0 \rightarrow$
2) stop

$m = 1$

Why Lasso Regression creates sparsity?

Home Insert Draw View Help

Shapes Ink to Shape Ink to Text

$$m = \frac{-100 + \lambda}{50}$$

$\lambda = 0 \quad m = -2$

$\lambda = 50 \quad m = -1$

$\lambda = 100 \quad m = 0 \rightarrow 1$

$\lambda = 150$

$m = 1$

$\lambda = 150, \; m = -5$

1) 0 →
2) stop

$$m = \frac{\sum (y_i - \bar{y})(x_i - \bar{x})}{\sum (x_i - \bar{x})^2 + \lambda}$$

0.1 → 0 ✗

$\lambda \rightarrow$ Ridge $<<0$ $==0$

Lasso → $\lambda \rightarrow$ numerator

$\lambda = 10000000$

Denominator